DOCUMENT RESUME

ED 155 193 TM 007 083

AUTHOR Hansen, Joe B.
TITLE Quality Control in Title I Evaluation: Problems and Approaches.
PUB DATE Mar 77
NOTE 16p.; Paper presented at the Annual Meeting of the American Educational Research Association (61st, New York, New York, April 4-8, 1977)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Administrative Principles; Compensatory Education Programs; Data Analysis; Evaluation Methods; Interagency Coordination; Interdistrict Policies; *Program Administration; *Program Evaluation; *Quality Control; Research Design; *Research Problems; School District Autonomy; Standards
IDENTIFIERS *Elementary Secondary Education Act Title I; *Meta Evaluation; Portland Public Schools OR

ABSTRACT

Educational program evalautiocn is subject to many difficulties. Problems with evaluation include such conceptual problems as: (1) understanding of decisions to be made; (2) adequate definition; (3) consensus on values and criteria; (4) coordination among administrative levels; and (5) use of an appropriate evaluation model. Technical problems include poor data collection, faulty analysis, and inappropriate inferences. Elementary and Secondary Education Act Title I programs have been closely regulated, but this has not eliminated the difficulties of evaluation. Therefore the Portland, Oregon Public Schools established, as a function of the central Evaluation Department, the responsibility of auditing the evalutions performed in each of the three sub-districts. Professional standards for evaluation data, methodology, and reporting were developed in cooperation with the sub-districts and published as formal guidelines. Although the auditing procedure was planned to consist of three stages, only two stages were actually completed: the review and critique of the design and instrumentation, and the review and critique of the final evaluation report. Time constraints prevented completion of the second stage, monitoring of the data collection activities. Suggestions for successful auditing are included; these guidelines describe personnel selection; effective communication; systematic standards; and administrative support. (CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED155193

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Joe B. Hansen

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

QUALITY CONTROL IN TITLE I EVALUATION:

Problems and Approaches

Joe B. Hansen
Northwest Regional Educational Laboratory
710 S.W. Second Avenue
Portland, Oregon 97204

MARCH 1977

Paper presented at the annual meeting of the American Educational Research Association, New York, New York, April 1977

TM007 083

QUALITY CONTROL IN TITLE I EVALUATION:

Problems and Approaches

Joe B. Hansen
Northwest Regional Educational Laboratory*

PART ONE: The Problem

Introduction

The purposes of this presentation are twofold: 1) to define and discuss some of the problems affecting the quality of evaluation data. 2) to review briefly, a procedure used in one decentralized school district to improve the quality of Title-I evaluation.

Factors Affecting Data Quality

Stufflebeam, et al., (1971) identified five major problems affecting the quality of evaluations, the problems of (1) definition, (2) decision making, (3) values and criteria, (4) administrative levels and (5) the research model. Each of these problems affects the quality of evaluation in a different way. The manner in which evaluation is defined establishes limits for the methodology of evaluation. For example, if evaluation is defined as measurement the methods used will focus on outcome data and the use of cognitive, behavioral and affective measures. This may result in the exclusion of many other types of evaluative data not amenable to measurement technology. The decision making problem is related to the use of the evaluative data. What decisions are to be made, by whom and through what processes are factors that affect the utility and interact with the quality of the data collected.

---

*Based on work performed while in the employ of Portland (Oregon) Public Schools.

Values and criteria also influence the quality of data in a profound manner. What are the value assumptions underlying the evaluation? Are they made explicit in the evaluation design and report? What criteria are used in making judgements about the effectiveness of the entity being evaluated?

Administrative levels pose a difficult problem in terms of the focus of and level of detail in an evaluation. Data useful to the decision making processes at building and project level are likely to be too fine grained for use at the SEA level. A means must be found for summarizing these data in a meaningful way to make them useful to state level decision makers. In like manner, state by state analyses are of little value to the federal agency responsible for administration of the over all program. The levels problems is easily recognized as one of the key problems leading to the development of a set of models for Title I evaluation, under section 151 of P. L. 93-380.

The fifth problem is that of equating evaluation methodology to research methodology. Some of the consequences discussed by Stufflebeam, et.al. include laboratory antisepsis (if it worked in the lab it will work in the classroom), experimental intervention versus a naturalistic approach, inapplicability of assumptions and impossibility of continuous refinement.

Horst, Tallmadge & Wood (1975) reviewed over two thousand Title I evaluations. Out of this total they identified only six in which the data were conclusive enough to state that the project either was or was not a success. In their analysis of the reasons for poor quality in the data, Horst, et. al. identified twelve common hazards to evaluation, the occurrence of any one of which could totally invalidate an evaluation. The hazards they identified were: use of grade equivalent scores, use

of gain scores, use of norm group comparisons with inappropriate test dates, use of inappropriate levels of tests, lack of pre and post test scores, use of non-comparable treatment and comparison groups, selection of project participants based on pretest scores, post-hoc matching to establish comparison group, careless administration and scoring of tests, faulty attribution, use of non comparable pretests and posttests, and use of inappropriate formulas to estimate posttest scores.

The five problems identified by the PDK group (Stufflebeam et al) and the twelve hazards identified by Horst, Tallmadge & Wood represent two different levels of interest in the problem of data quality. The PDK group, interested in the development of a conceptual model for evaluation, identified conceptual problems while the RMC group (Horst, et al) sought a more direct solution to the real problems facing Title I evaluation. Both analyses are useful and both deserve much greater attention than they can receive in this abbreviated format.

While all of these problems affect data quality they do not act in a consistent or predictable manner. Some are more salient in terms of their effects on Title I evaluation than others. Furthermore, any attempt to analyze the effects of such factors must recognize that they are not mutually independent, but are confounded in such a manner that they present a shifting kaleidoscopic effect on evaluation studies, the exact pattern of which is dependent upon the perspective from which it is viewed.

For example, the definition of evaluation that is appropriate will depend in part on the administrative level at which the data are to be used. In like manner the decision making process involves value assumptions and the employment of criteria. Levels, values and criteria are also interdependent, each influencing and shaping the others in an inexact and situationally defined manner.

The susceptibility of an evaluation study to one or more of the twelve hazards identified by Horst, et al, is a function of several factors, but evaluation expertise, quality standards, and evaluative accountability are perhaps three of the most important. In the ideal state, an evaluator with sufficient technical skill, operating under a clear set of quality standards with accountability for his/her evaluation as an established function, would be less likely to commit the errors identified as hazards than would an evaluator with less skill, no standards, no accountability.

It is obvious from the above that there are many problems that influence the quality of information collected and reported as evaluative data. The problems have both conceptual and practical implications and involve administrative as well as technical issues. Solutions to these problems must be sensitive to variations in standards, levels of expertise and evaluative accountability.

Problems Affecting Title-I Evaluation

Title I has been one of the most tightly regulated educational aid programs ever devised. It was the first to require outcome evaluation; but as we have learned, the mere requirement by the federal government in the absence of clear standards for evaluation quality did little to stimulate the production of useful information for U.S.O.E. to report to Congress on the effectiveness of the program nationally.

At state level the problems have been complex and varied. State Title I coordinators, lacking in authority and frequently untrained in evaluation skills have struggled valiantly to "compile data" from the

multitude of programs to produce reports that would meet the minimal requirements of the federal regulations. In practice, minimal requirements became maximal. The situation at state level, as was pointed out in Dr. Hunt's presentation has been fraught with a sense of hopelessness and despair proportionate to the conscientiousness of the state coordinator.

At the LEA level, where the data originate, things have not been so easy either. In the small districts where there is little evaluation expertise and no basis for a quality standard, the twelve common hazards are common practice. In the larger districts where there is some established expertise in evaluation and standards do exist, two kinds of problems emerge in the quality of evaluative data. One is the delegation of responsibility for Title programs and other non-contingent evaluations to relatively unskilled evaluators. Because Title-I funding is rarely dependent upon demonstrated program effectiveness the evaluation of Title-I is often taken lightly. It is therefore not uncommon for Title I evaluation to become the responsibility of the project director or Title I coordinator, whose expertise is likely to be in curriculum rather than evaluation. The second is a conflict in quality standards. This occurs in at least two ways: 1) the state level standard for evaluation quality is higher (more stringent) than that of the LEA; 2) the LEA standard is higher than that of the state. If the former is true the LEA may strive to raise the local standards to meet the state requirements. However, if the latter is true the LEA may be producing data that are more sophisticated than the state is prepared to deal with. In such cases data quality will be compromised by incorporation into the state report.

In decentralized districts, these problems of expertise and standards can be compounded by the interposition of another layer of administrative machinery, rules, standards and delegation of responsibility, such that

the standards for sub districts could be in conflict with one another and yet all could be in (minimal) compliance with the state guidelines. Such was the case in the Portland Public Schools three years ago. The result was that Title-I data produced by evaluators in each of the three, semi-autonomous areas of the district could not be interpreted in a common framework. Aggregation was not even to be considered. Three separate testing programs were producing three sets of independently normed data. Three different evaluators were using three different approaches to evaluate the Title I program in their respective areas.

Part Two: A Solution?

All of the problems outlined above led the Portland superintendent to establish as a function of the central Evaluation Department the auditing of evaluation work performed in each of the three areas (sub-districts). The purpose of the audit function was to establish quality standards and monitor the performance of the professional evaluators with respect to those standards. For practical reasons, Title I evaluation was chosen as a major focal point for the development of a model for establishing evaluative accountability.

Developing the Model: evaluation standards

It was recognized that in order to develop and effective evaluation audit a set of professional standards for evaluation data, methodology and reporting would be necessary. Such standards had been under development and were being refined through experience. The standards consisted of a set of procedures for developing and evaluation design: A Generalized Evaluation Design Procedure[1] and a set of Proposed Standards and Guidelines for Evaluation Reports[2]. The procedures and guidelines had received the

---

(1), (2) Unpublished, Portland Public Schools Evaluation Department

benefit of full review and criticism from the evaluation specialists in the district and had been generally adopted as a set of standards.

A procedure was developed for systematic review of each Title I evaluation from design to final report. The procedure consisted of a three stage review and critique process as shown in Figure 1.

The model did not operate exactly as intended. Therefore I will first describe the manner in which it was intended to work, after which I will discuss the departures from the intended plan, the reasons for the departures and their implications.

Stage one, the design audit was to consist of a thorough review and critique of the evaluation design, followed by a discussion between auditor and evaluator of the design features and auditor recommendations. This was intended to be a negotiating session, wherein design deficiencies and questions relating to instrumentation and planned data collection activities could be related to the evaluator's experience with Title I, knowledge of the student population and other factors to yield results that would meet the standards. The audit report and the evaluator's written reaction to it could be filed for later use in the implementation and evaluation report stages.

Stage two, the monitoring phase, was intended to include auditor site visits to project classrooms on a sampling basis, during data collection activities, to monitor all data collection. In this phase observations would be made regarding data collection procedures and any implications for data quality would be noted. The interim report would contain a summary of observations made and recommendations for procedural changes based on those observations. The evaluator would then have an opportunity to review and react to the recommendations. The intended recipients of the interim report would include both evaluator and program director.

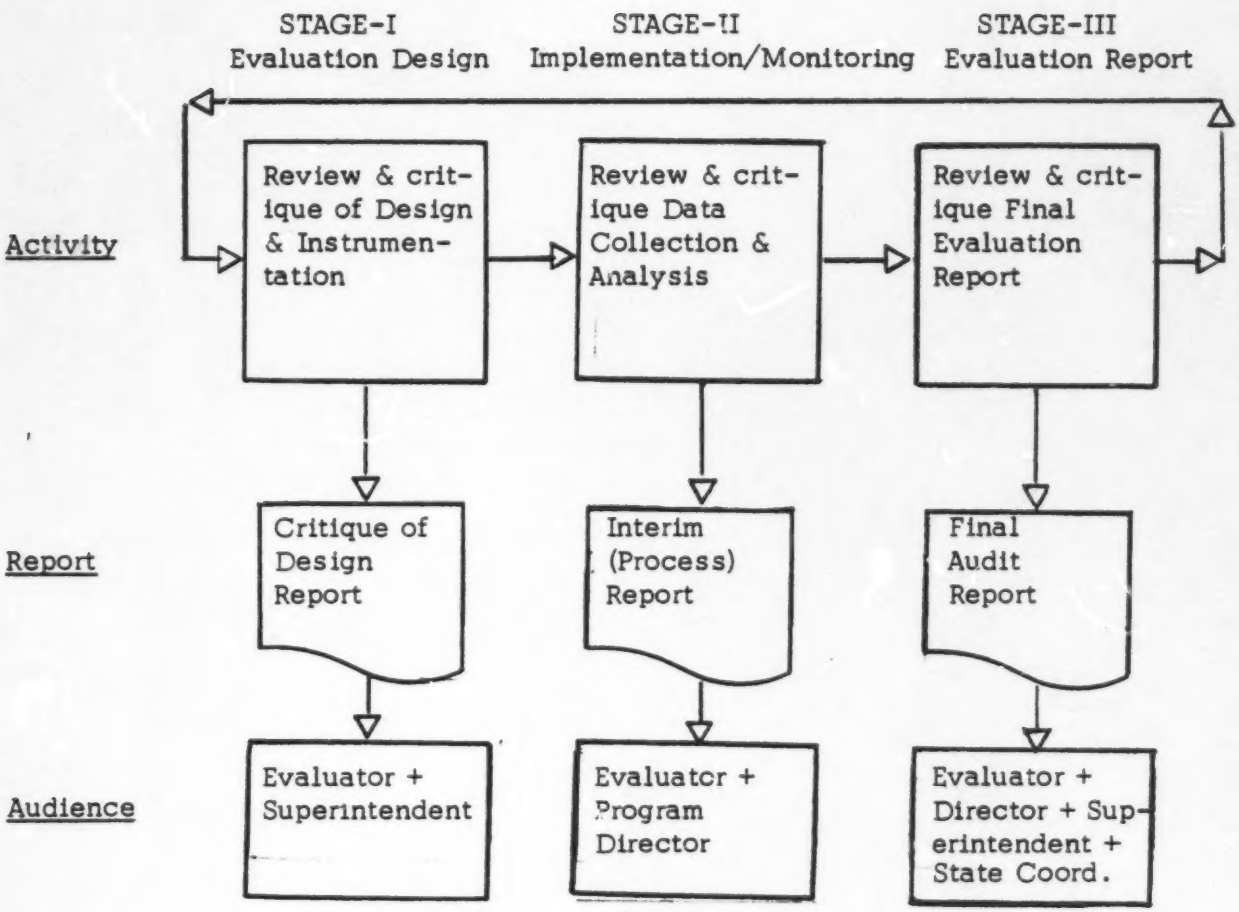

Figure 1.

Flow diagram of audit procedure

In stage three the final evaluation report would be critically reviewed according to the established standards and guidelines. The report would be examined for evidence of design adjustments and other responses to previous recommendations. The emphasis in this stage is on the appropriateness of data analysis techniques, the soundness of the conclusions drawn from the data and adherence to quality standards in report writing. The final evaluation report is the annual report sent to the state education office while the final audit report is distributed to the evaluator, the district superintendent and finally to the state compensatory education office.

At each stage of the model, the evaluator is given an opportunity to review and react to the findings and recommendations of the auditor. If possible, the evaluator and auditor should meet and discuss the audit findings thoroughly before the report for a given stage is finalized. In some cases telephone discussions and memos can suffice. Written reactions from the evaluator are encouraged and are attached to the reports prior to distribution.

Problems with the Model

Two major problems were encountered in applying the audit model to Title I. First was the problem of timing the application of the model to the Title I planning cycle. The decision was made to implement the model in the middle of the 1975-76 school year. This was too early for the final evaluation data to be available so it was decided that the audit would begin with a belated stage one analysis of the 1975-76 evaluation design, following up with stages two and three in the spring and summer of 1976. This choice of entry points was logical but not astute due to our inaccuracy in estimating the amount of effort required for a thorough analysis of the evaluation at each stage. The result was that by the time the first stage reports were completed there was no time remaining for the interim analysis and reports before it was time to begin the third stage reports.

The second major problem encountered in implementing the model was the stress placed on colleagial relationships by having a district evaluation specialist reviewing and critiquing the work of other evaluators. We had anticipated that some difficulties would occur and were therefore not surprised. Nonetheless, the problem was serious in terms of the effect on working relationships, especially between the auditor and one of the evaluators whose reports were found to be deficient.

Effectiveness of the Audit Procedure

As a result of these implementation problems some changes were made in the conceptual design and operating procedures of the audit model. First it was decided that the onsite monitoring and interim report (Stage II) were non essential aspects of the audit. The additional burden to the auditor, who had numerous other project and supervisory responsibilities, posed by this interim phase, was felt to be excessive when weighed against the value of the data generated. Furthermore, it was felt that this "looking over the shoulder" would prove to be irritating to the professional evaluators charged with conducting the Title I evaluations. This interim phase was therefore dropped from the Title I audit but was retained as an optional feature for other types of programs.

Second, as a result of the strain placed on colleagial relationships, an increased effort was mounted to inform each district evaluator of the purpose of the audit and the procedures employed. This included a greater emphasis on personal contacts by the auditor to discuss questions in depth with the evaluator prior to writing a report.

In spite of the difficulties mentioned, the audit-critique procedure proved to be effective in improving the quality of Title I evaluation and reporting. Evaluators in two of the three areas were responsive to the suggestions of the auditor and substantial improvements were made in the design, analysis and reporting of Title I evaluations. Specific improvements

were noted in statements of objectives, data analysis, data display, descriptive reporting and interpretation of data. The evaluator in the third area took umbrage at the criticism and refused to accept the validity of the audit.

A residual benefit that offsets the detrimental effect of the audit on colleagueship was that the auditor improved his professional relationships with the other Title I evaluators through the discussions that resulted from technical issues. Some of these issues led to long and stimulating conferences and telephone conversations over alternative methods of analyses. Many of the ideas generated in these discussions would show up later in the evaluation reports.

How to Make an Audit System Work

This form of audit is somewhat crude and some problems occurred in its implementation. However, an internal audit system can be an effective means of improving the evaluation of Title I or any other type of project provided that it is implemented correctly and clear guidelines are established for its use. The following suggestions are offered for those who might consider adopting this approach to evaluation quality control.

1. Personnel Selection:

Qualities required in an effective auditor include technical skill, thoroughness, precision, compulsiveness, diplomacy, humility and previously earned professional respect. He or she must be able to communicate effectively on technical as well as philosophical matters without alienating the evaluator or administrator. Evaluators as we all know, don't like to be criticized. They are used to being the paid critics and are likely to show resentment toward other evaluators who criticize them, especially as a matter of record.

2. Communications:

Effective communications about the audit procedure before its implementation as well as during its operation is critical to its success. Everyone who will be involved with it in any way, from the Board of Education to the state coordinator should be informed of the rationale, purpose, methods of operation and any other available information. Meetings should be held between evaluators and auditor-designee before the audit begins, to encourage the free exchange of ideas regarding how to make the audit most effective.

3. Standards:

The auditor will need a supporting system of evaluation standards to provide structure for the audit. The following elements are suggested as providing a skeletal basis for such a set of standards.

a. Definition of evaluation: A clear, meaningful definition of evaluation that is acceptable to the system and has been acknowledged by the evaluators and administrators whose work will be audited.

b. Evaluation design standards: Each evaluator whose work is subject to audit should be fuliy aware of the principles of evaluation design employed by the auditor in ascertaining the adequacy of a design. These should be spelled out in detail and distributed to the evaluators for reference use.

c. Methodological standards: Evaluators within the system should be aware of acceptable and unacceptable conventions, e.g., the twelve evaluation hazards. If there are preferred instruments for measuring particular outcomes, they should be made known. If the district has for one reason or another preferred and non-preferred methods of analysis, these should be made known along with their rationale.

d. Reporting standards: Evaluators should be informed regarding what is expected to be included between the covers of the evaluation report and how to format it. Standard format and reporting guidelines can greatly facilitate the work of the auditor and make reports more intelligible for all users.

4. Support:

Finally, the auditor must have the support of the system he/she is trying to improve. The role of an internal auditor is a discouraging role, with few intrinsic rewards. It takes courage and ego strength to perform well in the job. Take your auditor out to lunch, buy him a drink, tell him a joke, invite him to your house. You might even learn to trust him, if you try.

These are just a few suggestions. I hope you find them useful. It may be worth noting that evaluation quality within a district could be improved in most cases without the use of an audit if evaluation design, procedure, report content, and communications requirements were standardized and clearly communicated to evaluators.

REFERENCES

1. Horst, D. P., Tallmadge, K. T., Wood, C. T., A Practical Guide to Measuring Project Impact On Student Achievement, U. S. Government Printing Office, Washington, D. C., 1975.

2. Stufflebeam, D. S.; Foley, W. J.; Gephart, W. J.; Guba, E. G.; Hammond, R. L.; Merriman, H. O.; Provus, M. M., Phi Delta Kappa National Study Committee on Educational Evaluation and Decision Making, Peacock Itasca, Ill., 1971.